WEBSITE INTEGRITY AUDIT TOOL

[1]N.G.Bhanupriya, [2]D.Dhivya, [3]A.Gurusri, [4]Dr.R.Ramesh
[1,2,3]UG Students, Department of Computer Science and Engineering, RMK Engineering College, Chennai, INDIA
[4]Scientist-E, Department of Electronics and Information Technology, National Informatics Centre, STPI, Chennai, INDIA

Abstract— The Private Organizations host several websites to serve the customers in transparent, efficient and in a time bound manner. Some websites are hosted for their internal purposes also. In such cases, the sensitive data being accessed only within an organization are getting hacked. The websites are hacked in such a way that the organizations come to know only after a specific period of time. This project envisages to demonstrate how the Private organizations getting secured with the help of web base Audit Tool which is an OSS tool and gets updated with all the registered websites in a routine manner.The project scope involves the complete monitoring of the various websites being used by the private company. The user will be given a SMS or E-mail alert when some changes are made to the website.

Index Terms— ***data, hacked, secured, websites, updated, hash value, routine, monitoring***

I. INTRODUCTION

A secured web based monitoring audit tool developed using cryptographic algorithm namely SHA-256. This tool checks the integrity of the websites of any private organization. This tool gives a status report regarding the website integrity for any specific time given by the user. Snapshot Capturing, Updating the hash values, Integrity Check and Alerting the users are some of the main features of this tool.

II. OVERVIEW

The overall system architecture of the web application has 4 components which are listed below

- User Registration
- Audit Tool
- Metadata Updation
- Alerting System

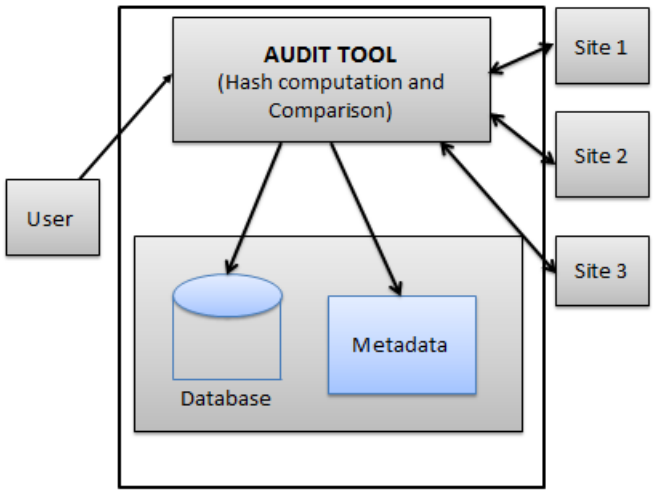


Fig.1. Stystem Architecture

.

A. User Registration

The user in the Fig.1 system architecture is the person in charge of monitoring all the websites of the company. Initially, the user registers with the Audit tool and with company name, websites, login id, mobile number, email id. The registered users will be given a Licence Id and Password to login to the Audit tool.Units

B. Audit Tool

This is the main part of this architecture where hash computation and comparison of hash values take place. Initially, the tool takes a snapshot of the website to be monitored. Then it computes the hash value for all the web

pages using SHA-256 Algorithm. The generated hash values are checked with the values that are previously stored in the database. This process is repeatedly done for all web pages including the image files stored in it.

1) FEATURES OF AUDIT TOOL

- It is more secure and provides user friendly interface.
- The tool is compatible with any platform.
- Users can fix a particular time interval for monitoring the websites.
- Example: The user of ABC Company can fix a time interval of 5 minutes to audit only few sensitive webpages of a website.
- A status report is generated regarding the website.
- The tool has an alerting system which will alert the user regarding any security problems with the websites.

C. Meta Data Updation

The hash values of the monitored website are stored in the base table. The meta data information of that base table is made available in the siteurl table.

TABLE.1 SITEURL TABLE

Webpages	Hash Value
Index Page	c256697f753f2edca443e9180fb2f8c13569f27fb 030f146c78616ba704a002e
About Us Page	40585558fe3085dc50faffb88c13023573b0bf38 e8d3b6655ebd21c2a6e031d8
Contact Us Page	78d3260a211ea1ac525e315619654549f2defc01 c44332b3077d41d671077b07
Services Page	dffde01b3818bb16d5b01727a27317b84953edc acff430929f2ce87514e3d3d1

This table (TABLE.1) stores the URL of all the monitored websites along with the time and status of last auditing. This table also gives the corresponding hash values of the website. The website is monitored for every 5 minutes and the status is updated in database.

D. Website Monitoring

This component in system architecture diagram illustrates that the website of a Private Organization are hosted by other server. Hence the data is retrieved and stored in a file. Then, the computation of the hash value for the website take place. This will not affect the security risk with the third party who has hosted this website.

III. SHA-256 ALGORITHM

SHA-256 operates in the manner of MD4, MD5, and SHA-1: The message to be hashed is first

(1) padded with its length in such a way that the result is a multiple of 512 bits long, and then
(2) parsed into 512-bit message blocks $M^{(1)},M^{(2)},M^{(3)},\ldots$
The message blocks are processed one at a time: Beginning with a fixed initial hash value $H^{(0)}$, sequentially compute

$$H^{(i)}=H^{(i-1)}+C_M^{(i)}(H^{(i-1)}),$$

Where C is the Sha-256 compression function and + means word-wise mod 2^{32} addition. $H^{(N)}$ is the hash of N.

A. Explanation

The Secure Hash Algorithm operates in such a way that a 256 bit hex values are generated for every content of a file, image etc. The hex value generated will be unique for every file. Thus changes made in any part of the file will be identified easily. The SHA-256 compression function operates on a 512-bit message block and a 256-bit intermediate hash value. It is essentially a 256-bit block cipher algorithm which encrypts the intermediate hash value using the message block as key.[6] Hence there are two main components to describe:

(1) The SHA-256 compression function

(2) The SHA-256 message schedule.

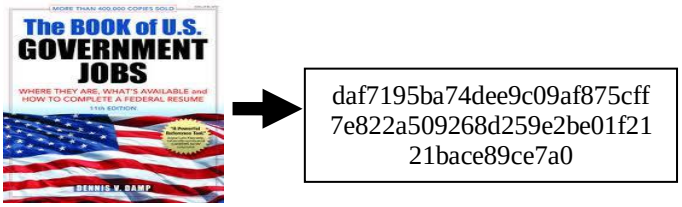


Fig.2 SHA-256 Image Hashing

The above Fig.1 SHA-256 image hashing shows the conversion of an image to a 256 bit hash value. The SHA-256 hashing for a html file is given below

```
//Sample.html

<html>

<body>

Hello World !!!!

</body>
```

</html>

The generated hash value for the above html file is

e8953b5b1e4e406c1b188081de377a25295e07228f8770664d20ff7dfeb2a85e

IV. MODULE DESCRIPTION

A. .Initial Snapshot

The first module illustrates about capturing snapshot of the website to be monitored. Initially during registration with this audit tool, the User will be given a valid license id. The user has to enter the valid id and the website to audit. This will be done with the help of Apache Tomcat Server and the authentication using JSP.

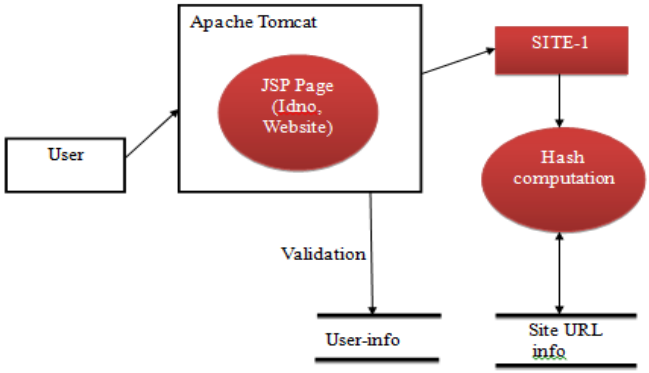


Fig.2 Initial Snapshot

The entered id will be checked with the database and if matched then the user is a valid user else it will be redirected to Login page. If it is a valid id, then the audit tool will capture a snapshot of the website entered. The snapshot is taken using SHA-256 cryptographic algorithm. This algorithm will read the complete the html contents of the website. After reading the html contents, a hash value will be generated for the corresponding web page.[4]

B. Database Updation

This module tells about updating of the database with the computed hash values. If the website is monitored for the first time, then the values are entered into the database. Else the computed hash value is compared with values previously stored in database. The audit tool executes the same for all the web pages of the entered website. The site is monitored and the hash values are updated in the database when it is monitored for the first time.

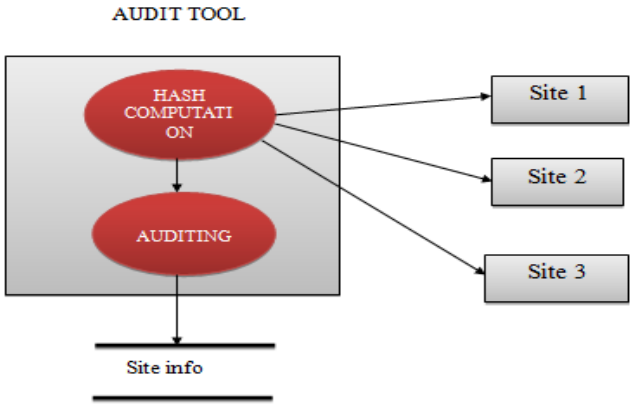


Fig.3 Database Updation

C. Batch Verification

This module explains the batch processing of audit tool. The audit tool will be monitoring the website for the time interval specified by the user. This tool will make the process run at the backend until the user interrupts. The purpose of this module is to make the process run without human intervention in the given time limit specified by the user. The user can even recheck the particular web pages when needed.[2]

D. Alerting System

This module describes the alerting process done when the website is modified. The computed hash value is checked with the previously computed values in the database. If they match, then the website is not hacked and the auditing can be done for all similar websites. If they don't match, then the website is hacked or modified and the alert function is called. This function sends an e-mail or SMS alert to the valid user who monitors the website.

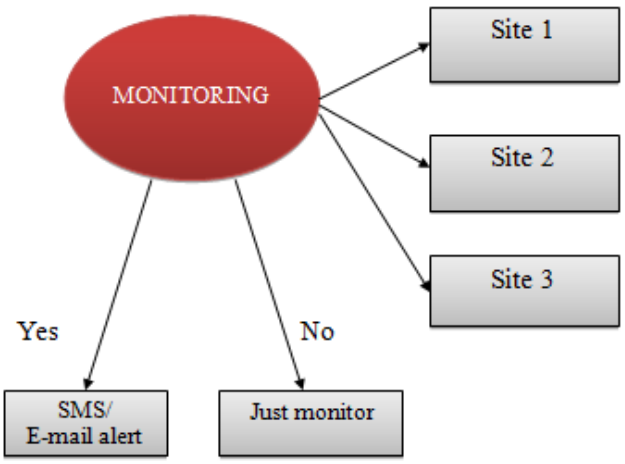


Fig.4 Alerting System

V. IMPLEMENTATION

The implementation stage involves careful planning, investigation of the existing system and methods to achieve the result of the project. Implementation in this project focuses on private organization registration with the audit tool and monitoring the list of websites as needed. The audit tool will visit the website to be monitored and will capture the snapshot of that website. The tool will then compute the hash value and update those values to database. The updated value is then checked with the previously computed hash value. If both the values are same, then the audit tool will continue to monitor the website, else it will raise alert to the user via E-mail/SMS. The critical factor that is to be analyzed is that the audit tool will give an alert only to the user who is responsible for monitoring all the website of that particular private organization. Thus the information of that user is obtained during the registration of the private organization with this tool.

The implementation can be preceded through any Operating System as this application is purely web based. The application is compatible with any browser and can be best viewed in Google Chrome, IE 7 and above versions. The database used for this application is MySQL. The database maintains the user information details while registration and also the computed hash values when monitoring. The implementation phase deals with the issues of quality, performance, baselines, libraries, and debugging.

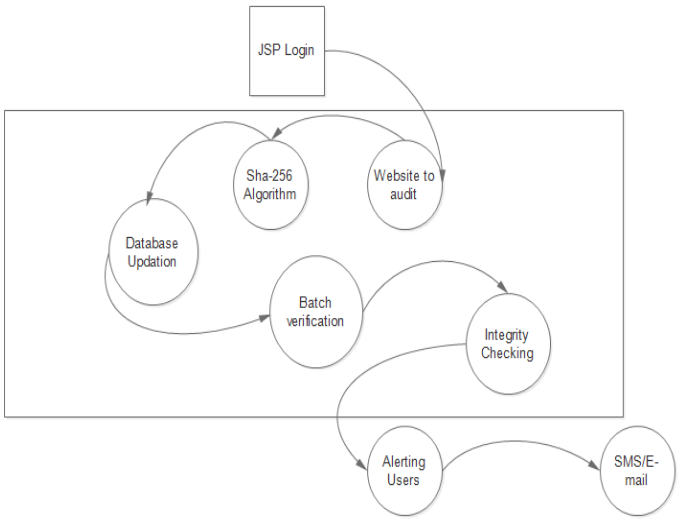


Fig.5 Data Flow Diagram

The data flow diagram in Fig.5 visualizes the control flow of various processes. Initially the user who logs in to the audit tool with registration id, enters the website to audit. The control is now transformed to the auditing website which is the feature of the audit tool. Then the audit tool captures the snapshot of the website and computes the hash values using SHA-256 Algorithm. This hash value is checked with previously computed hash value that are stored in the database. The batch processing of the application is done and the integrity of the website is checked for user defined time interval.

The status report is generated and the user will find the webpages that are modified. In addition to this, the batch processing of the application is done, where the user can get an alert via E-mail/SMS when a website is modified.

VI. PROCEDURE

Steps

1. Start the process
2. Go to the registration page and enter the company name, website, email id, mobile number and press submit.
   2.1 The login page will be redirected and now enter the license id, website to monitor.
3. Login id entered is checked with id stored in database using JSP which runs on Apache Tomcat Server.
4. If it is matched, then it is redirected to the website which is entered by user in login page
   4.1. Else it is redirected to Login page.
5. The Audit tool will read the site details and take a snapshot.
6. The hash value of the website is computed using SHA-256 hash algorithm.
7. The computed hash value is updated in the database.
8. Then compare the computed hash value with database hash value.
   8.1. If it is matched then the website is not modified.
   8.2. Else the website is modified and the alert is raised via e-mail or SMS.
9. Batch processing of the application is done and the site is monitored for every 5 minutes. Repeat step 5 to step 8.
10. Stop the process.

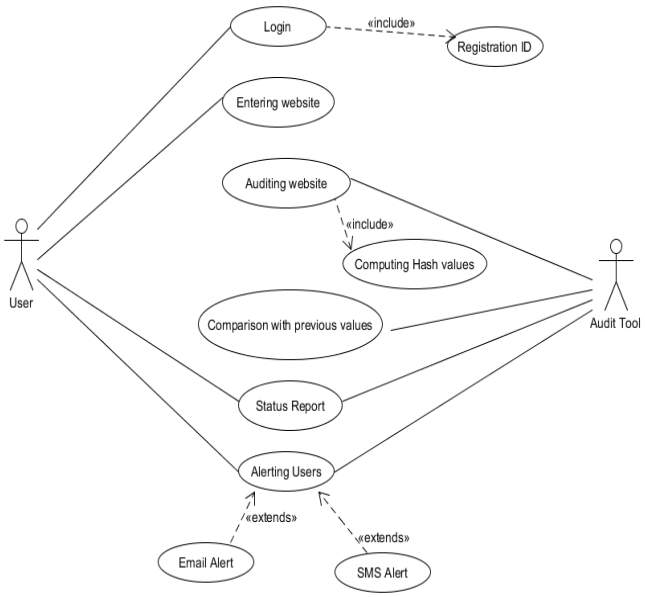


Fig.6 Use Case Diagram

The user in the use case diagram in Fig.6 depicts the functions of the user with the audit tool. The use initially has to register themselves with the essential details like company name, company website, email id, mobile number with this audit tool. The audit tool will then provide a registration id to the user. The user then enters the website name to audit. The audit tool takes a snapshot of the website and computes the hash value. This hash value is checked with the previously stored hash value in the database. If the comparison has failed, then an alert is raised via E-mail/SMS. The status information is updated to the database and displayed to the user.

VII. RESULTS AND DISCUSSION

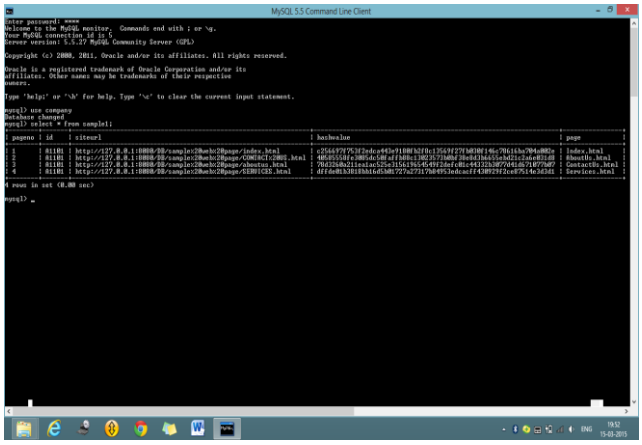
The results and discussion for every module is explained below.

A. Login Authentication

This is a snapshot of login authentication where the user enters the registration id, website. The entered id is checked in the database and an alert box is raised stating that the id is valid.

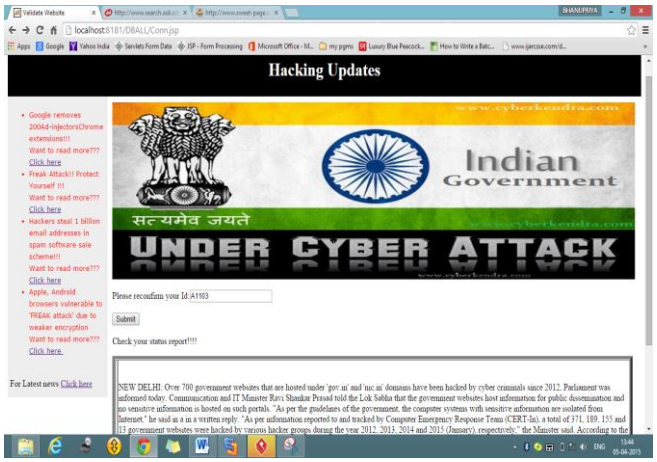
B. Site Information

The audit tool retrieves the site url from the site information table. The computed hash value is checked with table shown below.



C. Monitoring the Websites

The site is checked for every 5 minutes and the status is simultaneously updated in the database. The batch processing is done for every 5 minutes. This time interval can be changed according to user interest. The batch processing can be done for all the webpages of a company or only to some sensitive pages.



This page shows the current issues regarding website hacking in India. The sensitive factor is that over 700 government websites which are hosted under "gov.in" are being hacked.[5]

D. Status Report Generation

This is the status report that is generated for id (A1101) after the monitoring of the sample website. This report is generated for every 5 minutes and this time interval can be adjusted by the user according to their use. If a user wants to recheck the page, then the web page URL is clicked by the user and page is again rechecked by the user. The user whose details are stored in website will be given an alert if changes are made to the website. If the Organization has made some changes, then the user has to notify the tool about these changes. Else the tool will raise an alert to the user which might be ignored.

VIII. CONCLUSION

We have introduced a secured hash algorithm for integrity checking of the websites where a snapshot of the complete website including all the webpages, images, links are taken. The hash value for every snapshot is captured using SHA-256 algorithm and the hash values are updated in the database. These updated values are then checked with the previously stored hash values in the database. This comparison is done and a status report is generated to the user. This process is made as a batch verification where the website is monitored for every 5 minutes. An alert is raised via SMS or E-mail when any modification is done to the website without the knowledge of the user.

REFERENCES

[1] https://sucuri.net/services/web-integrity-monitoring

[2] http://oakland09.cs.virginia.edu/papers.html

[3] http://en.wikipedia.org/wiki/Web_application_security

[4] https://github.com/apache/jmeter

[5] http://www.ndtv.com/india-news/more-than-700-government-websites-hacked-since-2012-745898

[6] http://en.wikipedia.org/wiki/Secure_Hash_Algorithm